SAMGrid Status Report

Adam Lyon, 20 December 2005 GDM

See the last section for implications of a CDF Analysis Disk Pool

1 Project Drivers, Scope and Milestones

These are taken from the FY06 SAMGrid Budget Document. Updates are based on the November 2005 report.

1.1 Continuing Activities

1.1.1 Continue Smooth Operations

DB Server Improvements: Steve White has fixed a serious thread safety issue in the DB Server and has made minor improvements in speed. He is next going to tackle the python wrapping of CORBA transfer data-- this wrapping was added in v6 and appears to be the main source of non-performance.

1.1.2 Complete full deployment of SAM DH at CDF

- SAM on the farm is still using v6 (not the frozen v7). The CDF farm group needs to update their scripts and do testing. I understand they are manpower limited for this change, but it needs to be done.
- SAM is being integrated in their CAF restart: This is progressing slowly. In January I will make this a top priority for SAM – hopefully the CAF developers will do the same.
- CDF wants to be able to transfer a file out of SAM (e.g. a Root file) to any node running GridFTP. v5 SAMGrid already has a facility to do this that does not work in v7. Andrew has made updates to the v7 code and is in testing. A problem with the dCache gridFTP door was discovered – dCache experts are being consulted.

1.1.3 DØ MC & SAMGrid

Latest version of SAMGrid was brought to production.

1.1.4 LCG & OSG Integration

The SAMGrid/LCG forwarding gateway is in production. A proposal to handle job accounting for DØ is being circulated.

1.1.5 Integrate with VOMS/VOMRS

No news as Parag was away for much of last month.

1.1.6 SQLBuilder

Randolph is learning inner workings of the DBServer in order to integrate it with the SQLBuilder. We will have a meeting this week to finalize the changes to the dimensions language.

1.1.7 DØ upgrade from v5 to v7

Dehong has examined the v5 SAM DØ online code and it appears it will not be difficult to port the parts that are missing from v7. This is a major relief as the callback mechanism was removed in v7 and has to be put back into place for use by DØ online. It will not be hard to restore that functionality.

The DB population scripts need to be written to make use of the new v7 MC request system. This is in progress.

1.1.8 Rewrite broken groups and quotas for SAM managed cache

The human resources for this project have been retasked (Igor M.).

1.1.9 Deploy new SAM Data Handling Monitoring

A test version of "SamHDTV" is operational and working correctly. The MIS and mainline station codes were merged and final testing will begin before a test deployment.

1.1.10 Testing

All of the DB Server unit tests have been wrapped in our SAM Test Harness, making testing new releases much easier. We also need to make specific tests on the client in order to test the affects of Python 2.4. These tests are being written.

1.2 Moving forward with new technology (new activities)

1.2.1 Integrate SAMGrid with v6/7 compatible Run Job

Working to remove the application particulars of a job out of SAMGrid and put into RunJob. One problem has been Python versions. SAM uses an old version of Python (2.1) lacking many advanced features. We have recently tested our packaging ("freezing") system with the latest version (2.4) with much success. We are going to start deploying SAM software that uses the latest Python.

It was discovered that dcOracle (the python package used to interface the DBservers to Oracle) does not work in Python 2.4. dcOracle is no longer supported by the author (not at Fermilab). A different python to Oracle interface product, called *cx*, is supported and is supposed to be interface compatible with dcOracle (both use a standard Python-DB interface). Steve

White is investigating use of *cx* in the DBserver. Fortunately, the DB interface is abstracted in the DB server so a change should be easy.

1.2.2 Investigate deployment of SAM redundant information services

This project is Sinisa's Information Service system. No work has been performed in the past month.

1.2.3 Investigate deployment of SAM web services

MINOS has been testing Sinisa's SAM web services prototype. They provided feedback and some bugs have been fixed. The SAM team right now does not have the resources to pursue large scale production testing at the moment.

1.2.4 Investigate use of Enth for data base queries (continuation of SBIR project)

No work has been done. Awaiting Matt's report. Still no report.

1.3 Providing new capabilities

1.3.1 SAM DH and Condor Glide in

Initial discussions are beginning to determine requirements. No additional progress.

1.3.2 SAM Edge Service prototype

The Wisconsin student supposedly has working scripts to do the on the fly deployment, but I have seen no report yet.

1.3.3 SAM usage of SRM capable storage elements

We have begun planning and design of the SRM and SAM DH interface. We had a very fruitful meeting with Timur and now understand the current and future capabilities of the dCache SRM interface.

1.3.4 Implementation of SRM interface around SAM managed cache

No work has been performed yet.

1.3.5 Investigate breakup of SAM data handling services

No serious work has been performed yet.

1.3.6 Investigate SAMGrid for Analysis

No work has been performed yet.

2 Effort

Fermilab CD effort is ~6.0 FTE (as of the December Effort Reporting)

- 100%: Andrew, Parag, Valeria, Steve Sherwood
- 50%: Randolph, Adam, Steve White, Robert, Krzysztof, Dehong
- 20%: Gabriele

Note that Valeria's guest scientist position ends at the end of December.

Breakdown of effort is below. Note that time off (vacation, sick, holiday) is not included, so the total effort will not match the available effort.

Effort	FTE
Core Development	2.5
Deployment to Production	1.4
Operational Support	1.0
Project Management	0.5
Outreach	0.1
Total	5.5

3 Risks

The risks are unchanged from November.

Some of the previous risks (unreasonable expectations and feature creep) are somewhat under control as we are now bringing related requests to the GDM instead of handling ourselves. A lesson that I'm learning is to always insist on use cases and requirements before any further consideration is made on a request.

Some new risks...

- Human resources: While I think we are in ok shape now to handle the projects we've started, we do not have the resources to start other important projects (Web services deployment, breaking up SAM into services). We are also undertaking some short term rapid projects to get some operational problems out of the way (speed up DB server). I am hoping that completing the CAF restart, monitoring, and DØ migration will free us up a bit.
    - The loss of Valeria will be deeply felt. She has been indispensable in CDF SAM operations and development of some CDF specific SAM components (e.g. their framework interface to SAM). She has also been a vocal supporter of SAM and often a voice of reason.

- Upgrade to Oracle 10g from 9i. It could be great - or a disaster. The speed of many SAM queries worsened when we switched from 8 to 9.

We need a good testing program to find problems before 10g goes into production. MINOS has done some preliminary testing and found no problems - but their database is a small fraction of the size compared to CDF and DØ.

- Grid politics – I understand that protocols and interfaces may be changed (e.g. Condor using their own protocols) from what we use currently. SAMGrid will need to keep up.

4 Implications of a CDF Analysis Disk Pool

It is difficult to understand what CDF will require of SAM Data Handling without knowing the use cases. As of yet I have not seen use cases for this facility. But here are some guesses as to what they will need.

- We already have a dCache interface SAM station which is in use for the main CAF dCache. The station can only handle one dCache instance, therefore a different station will need to be set up for the analysis disk pool.

- We never anticipated having more than one dCache at CDF. Will they share the same /pnfs hierarchy? That is will there be some way to distinguish the location of a file (whether it lives on the main dCache or the analysis dCache)? If so, then the stations can be configured to deliver URLs from their particular dCache. If there is no way to distinguish the location, then there will be no way to tell the analysis disk pool station to not deliver files it doesn't have (and since there is no tape back end, deliveries will fail). The user will have to ensure that he or she only asks for files that are pinned to the analysis dCache -- SAM cannot do this for them.

- Is SAM going to be storing files to the analysis disk pool (e.g. as a durable cache)? I'm not sure if we can store directly to a tapeless dCache (now, we store to enstore and the file gets staged to dCache when requested or manually pinned).

- Because the analysis pool would need its own SAM station, users on CAF would have to specifically choose that SAM station.

- If the main use case is to access files from the desktop (not via a batch system), then the implementation of SAM needs to be determined. SAM can be used as merely a file catalog and files can be requested directly from dCache. But then there is no management and throttling of the dCache resource, and overloading of dCache may occur. If files

are requested through SAM, then there will be a latency as SAM throttles usage to prevent overloading dCache.